HP StorageWorks

Continuous Access EVA planning guide

Part number: AA–RW1DA–TE
First edition: December 2004

invent

Continuous Access EVA planning guide

Contents

Figures

Tables

About this guide

This guide provides information about HP Continuous Access EVA:

- Basic concepts
- Array configurations
- Fabric configurations
- Solution planning

Intended audience

This guide is intended for Information Technology (IT) managers, business managers, and SAN architects who are responsible for designing, evaluating, and selecting replication solutions.

Prerequisites

Prerequisites for using this product include knowledge of:

- SAN fabric configurations
- Disaster planning
- HP StorageWorks Enterprise Virtual Array (EVA)

Related documentation

In addition to this guide, please refer to other documents for this product:

- *HP StorageWorks Continuous Access EVA administrator guide*
- *HP StorageWorks Continuous Access EVA overview*
- *HP StorageWorks Continuous Access EVA performance estimator user guide*
- *HP StorageWorks Continuous Access and Data Replication Manager SAN extensions reference guide*
- *HP StorageWorks Command View EVA user guide*
- *HP StorageWorks EVA replication compatibility reference*
- *HP StorageWorks Enterprise Virtual Array user guide EVA3000*
- *HP StorageWorks Enterprise Virtual Array user guide EVA5000*
- *HP StorageWorks Replication Solutions Manager installation guide*
- *HP StorageWorks Replication Solutions Manager online help and user guide*
- *HP StorageWorks SAN design reference guide*

To find these documents, go to the web sites listed below.

- **Continuous Access**—
  http://h18006.www1.hp.com/products/storage/software/conaccesseva/index.html
- **Command View EVA**— http://h18006.www1.hp.com/products/storage/software/som/index.html
- **SAN design and SAN extensions**—http://www.hp.com/go/SANDesignGuide

Document conventions and symbols

Table 1 Document conventions

Convention	Element
Medium blue text: Figure 1	Cross-reference links and e-mail addresses
Medium blue, underlined text (http://www.hp.com)	Web site addresses
Bold font	• Key names • Text typed into a GUI element, such as into a box • GUI elements that are clicked or selected, such as menu and list items, buttons, and check boxes
Italics font	Text emphasis
`Monospace font`	• File and directory names • System output • Code • Text typed at the command-line
`Monospace, italic font`	• Code variables • Command-line variables
`Monospace, bold font`	Emphasis of file and directory names, system output, code, and text typed at the command-line

WARNING! Indicates that failure to follow directions could result in bodily harm or death.

CAUTION: Indicates that failure to follow directions could result in damage to equipment or data.

IMPORTANT: Provides clarifying information or specific instructions.

NOTE: Provides additional information.

TIP: Provides helpful hints and shortcuts.

HP technical support

Telephone numbers for worldwide technical support are listed on the HP web site: http://www.hp.com/support/.

Collect the following information before calling:

- Technical support registration number (if applicable)
- Product serial numbers
- Product model names and numbers
- Applicable error messages
- Operating system type and revision level
- Detailed, specific questions

For continuous quality improvement, calls may be recorded or monitored.

HP strongly recommends that customers sign-up online using the Subscriber's choice web site at http://www.hp.com/go/e-updates.

- Subscribing to this service provides you with email updates on the latest product enhancements, newest versions of drivers, and firmware documentation updates as well as instant access to numerous other product resources.
- After signing-up, you can quickly locate your products by selecting **Business support** and then **Storage** under Product Category.

HP-authorized reseller

For the name of your nearest HP-authorized reseller:

- In the United States, call 1-800-345-1518.
- Elsewhere, visit http://www.hp.com and click **Contact HP** to find locations and telephone numbers.

Helpful web sites

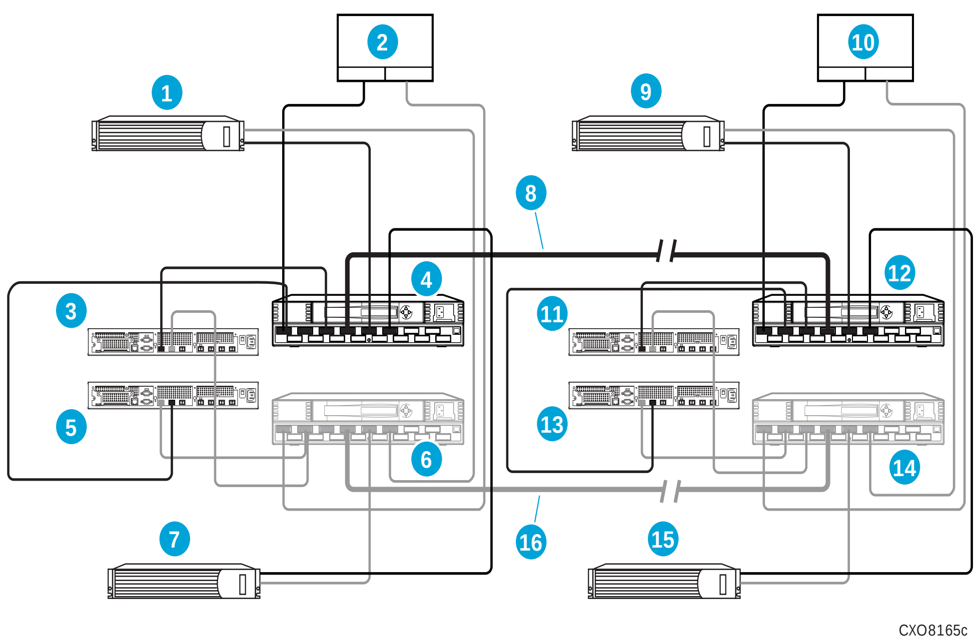
For third-party product information, see the following vendor web sites:

- http://www.hp.com
- http://www.hp.com/go/storage
- http://www.hp.com/support/
- http://www.docs.hp.com

1 HP Continuous Access EVA overview

HP StorageWorks Continuous Access EVA is the remote replication component of HP StorageWorks Enterprise Virtual Array (EVA) controller software. When this component is licensed, the controller copies data online, in real time, to a remote array over a a local or extended storage area network (SAN). Properly configured, HP Continuous Access EVA is a disaster-tolerant storage solution that guarantees data integrity if an array or site fails.

Figure 1 shows a basic configuration with redundant arrays and fabrics. One array is located at a local (or active) site and the other at a remote (or standby) site. In the figure, one fabric is called the black fabric and the other is called the gray fabric. Each array can perform primary data processing functions as a source, with data replication occurring on the destination array. The replication process can also be bidirectional, with some I/O streams moving to the array and other I/O streams moving simultaneously from the array. This feature allows the array to be the source for some data groups and the destination for others.



1 Local active management server	**2** Local host
3 Local controller 1	**4** Local black fabric switch
5 Local controller 2	**6** Local gray fabric switch
7 Local standby management server	**8** Interswitch link–black fabric
9 Remote standby management server 1	**10** Remote host
11 Remote controller 1	**12** Remote black fabric switch
13 Remote controller 2	**14** Remote gray fabric switch
15 Remote standby management server 2	**16** Interswitch link–gray fabric

Figure 1 Basic HP Continuous Access EVA configuration with redundant servers

In Figure 1, the management server represents the server where the EVA management software is installed, including HP StorageWorks Command View EVA and HP StorageWorks Replication Solutions Manager. HP recommends at least two management servers at each site to avoid a single point of failure. If a significant failure occurs at the source array location, redundancy allows data processing to quickly resume at the destination array. This process is called failover. When the cause of the array failure has been resolved, data processing can be moved back to the original source array by performing another failover. Also inFigure 1, the host represents any server that is using storage space on the array. It is also a server running applications that are not used for EVA management (such as Microsoft Exchange).

NOTE: For more information about HP Continuous Access EVA features, see the *HP StorageWorks Continuous Access EVA administrator guide*.

2 Designing a remote replication solution

This chapter describes business requirements and how they affect the design of a remote replication solution.

Topics include:

Determining business requirements

Identify the business requirements driving the need for an HP Continuous Access EVA solution. Common business requirements to consider are:

- High availability
- Disaster tolerance
- Recovery point objective
- Recovery time objective

High availability

The combination of redundant systems, software, and IT processes, with no single point of failure (NSPOF), reduces the risk of downtime and ensures high availability. Continuous Access EVA provides highly available and reliable access to *data*. However, it is a storage solution and does not provide highly available *applications*.

If you require highly available *applications*, you must include additional servers to provide application processing platforms if the primary server fails. For example, you can deploy a cluster of servers at the local site, with either a single backup server or a cluster of servers at the remote site.

Disaster tolerance

Disaster tolerance is the high-availability technology and services that enable the continued operation of critical applications during a site disaster. For example, if two sites are separated by a distance greater than the potential size of a disaster, then one site should be able to continue processing after the disaster. Continuous Access EVA enables applications to automatically and continuously build two copies of application data at separate sites that are far enough apart so that a single event does not destroy both sites.

NOTE: HP Continuous Access alone does not provide high availability and disaster tolerance. Both requirements are only a part of the business continuity model; you must consider the entire model if you are designing a business continuity solution. For more information, go to hp.com/go/continuity.

Recovery point objective

The recovery point objective (RPO) is the amount of data that can be lost at the time of the disaster. RPO is measured in time and typically ranges from zero to several hours. For example, with a short RPO, new data much be flushed frequently from the server cache into storage to prevent loss due to a server crash. A shorter RPO increases the need for real-time (synchronous) data replication.

Recovery time objective

The recovery time objective (RTO) is the length of time it takes to return an application to operation. With HP Continuous Access EVA, it includes the time to detect the disaster, fail over the storage, and restart the application on a new server. RTO is usually measured in minutes or hours, and occasionally, days. A shorter RTO increases the need for communications between the application and the storage.

Determining the threat radius

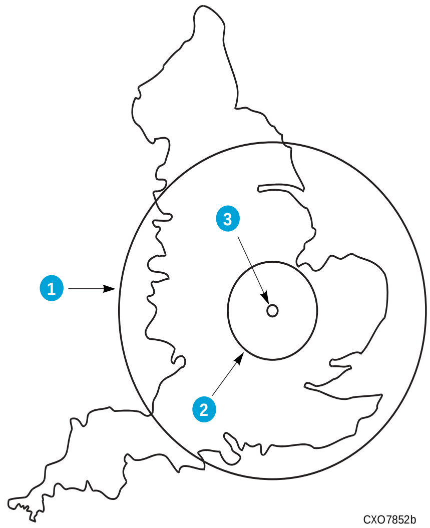
The threat radius is the distance from the center of a threat to the outside perimeter of that threat. For example, half the diameter of a hurricane or typhoon is the threat radius of that storm. The threat radius of toxic chemicals is defined by the strength of the wind, with the downwind threat radius much larger than the upwind threat radius, producing a more elliptical threat area.

The threat radius types are:

- **Local**—The threat is less than a few kilometers in radius (less than 25 square kilometers or 15.5 square miles). Local replication has the least effect on performance compared to the other options. Examples of local threats include tornados, fires, floods, and power loss.
- **Metropolitan**—The threat extends from 25 square kilometers to 5,000 square kilometers (3100 square miles) in radius. The performance impact due to replication beyond metropolitan-sized threats is similar in performance cost to running older disk drives—it is slower, but acceptable. Examples of metropolitan threats include large chemical incidents, moderate earthquakes, and severe storms.
- **Regional**—The threat affects a radius of hundreds of kilometers to tens of thousands of kilometers. A regional disaster requires the largest separation distance when planning disaster-tolerant solutions. Depending on the distances, data replication beyond a regional disaster threat radius affects system performance. For example, separation distances greater than 1,000 kilometers increase the cost of the link and slow down performance rather than provide disaster tolerance. Fortunately, these distances are rarely needed. Examples of regional threats include large floods, hurricanes, and typhoons.

Some threats may span two or more types, depending on size and severity. If this is a possibility for your environment, develop your solution based on the larger of the two threats.

Figure 2 illustrates the relative relationship between the threat types.



1 Regional (> 5,000 square km)

2 Metropolitan (up to 5,000 square km)

3 Local (< 25 square km)

Figure 2 Threat radius

When determining the threat radius, identify the threats to the source system, and if they apply to the backup system. For example, do not place both sites in the same flood plain because one flood could destroy both sites. Similarly, if severe storms tend to travel in a certain direction, then place the second site perpendicular to the expected route of travel, and as far apart as needed to prevent one storm from affecting both sites.

Consider any local regulatory requirements that can increase or limit the separation distance. For example, certain counties in the United States require both sites to remain within the same 100- to 400-square-kilometer (60- to 250-square-mile) county. This restriction has limited the maximum separation distance to less than 30 km (19 miles) in an area prone to earthquakes. Such earthquakes have affected buildings several hundred kilometers from the earthquake epicenter.

Measuring distance

The effect of distance on replication time does not change the type or scope of a potential disaster. This section describes the following topics:

- Intersite network requirements, page 16
- Defining latency, page 16
- Calculating latency, page 17

Intersite network requirements

Table 2 lists the HP Continuous Access EVA intersite network requirements. For more information about selecting the appropriate SAN extension, see the *HP StorageWorks Continuous Access and Data Replication Manager SAN extensions reference guide*.

Table 2 Intersite network requirements

Component	Requirement
Bandwidth	Dedicated to storage
Maximum transmission unit (MTU) of the IP network	1500 bytes
Maximum latency	**One way**—100 milliseconds **Round trip**—200 milliseconds
Average packet loss ratio	**Low loss**—0.0012% averaged over 24 hours **High loss**—0.2% averaged over 24 hours, not to exceed 0.5% for more than 5 minutes within a 2 hour window[1]
Latency jitter[2]	Not to exceed 10 milliseconds over 24 hours

1. A network that supports a storage interconnect or remote replication cannot have a packet loss ratio exceeding an average of 0.01% over 24 hours.
2. A measure of how stable or predictable the delay is in the network. It is the difference between the minimum and maximum values. The greater the jitter, the greater the variance in the delay, lowering the predictability of performance.

Defining latency

Latency is also called one-way delay, which is the time needed for the bits of a read or write to move from one site to another. The speed of light in a vacuum, for example, is approximately 3×10^8 meters per second or 186,000 miles per second. In fiber optic cables, this slows to about two-thirds of the speed of light in a vacuum, or approximately 2×10^8 meters per second. Converting the speed of light in fiber from meters per second to seconds per meter, the result is 5 nanoseconds per meter, or 5 microseconds per kilometer. The maximum Continuous Access EVA separation is limited to a one-way delay of 100 milliseconds, which is equivalent to a cable distance of 20,000 kilometers (approximately 12,500 miles), assuming no other delays. Table 3 lists other examples of one-way delays.

Table 3 Examples of one-way delays

One-way delay (ms)	Point-to-point cable distance in km (miles)
1	200 (125)
3	600 (375)
9	1,800 (1,125)
18	3,600 (2,250)
36	7,200 (4,500)
60	12,000 (7,500

Table 3 Examples of one-way delays

One-way delay (ms)	Point-to-point cable distance in km (miles)
100	20,000 (12,500)

NOTE: The one-way delay measurement for routed networks is longer than the equivalent for point-to-point networks, due to additional routing delays and the indirect paths needed to make the connection.

Calculating latency

Intersite latency is the distance (measured in time) between two sites. To determine intersite latency, use one of the following methods:

- **Network latency**—Use if an intersite network exists
- **Driving distance**—Use if an intersite network does not exist

TIP: If you are converting a point-to-point dedicated network to a shared routed network, allow 50% of additional latency (using either method).

Network latency

Ask the network engineers for an estimate of the one-way or roundtrip intersite latency. For example, the network engineers report that the current point-to-point network has a 24-hour average roundtrip latency of 2 milliseconds. If you are deploying HP Continuous Access EVA using a new routed network, use 3 milliseconds as the initial estimate.

The Internet protocol `ping` utility reports roundtrip latency. For example, using the management server, "ping" the internet router at the remote site as follows:

```
ping full-address-of-remote-router -n 3600 -l 2048
```

where:

- n is the number of pings to perform at one per second (3,600 seconds is one hour)
- `l` (lowercase letter L) is the size of the data packet (a Fibre Channel frame is 2,048 bytes)

Driving distance

To estimate the cable distance between two sites without a network, measure the distance by driving a car between the two sites. For point-to-point networks, multiply the driving distance by 1.5. For routed networks, multiply the driving distance by 2.25 to accommodate additional delays from network routing. Multiply either sum by five milliseconds to obtain network latency.

For a more accurate estimate, contact various network vendors and ask about possible routing and delays between the two sites in question. If there is a known delay, use it. If not, use the driving distance along the network path, and not the shortest distance between the two points.

Comparing the results

If possible, compare the results of both methods. If the current network latency is more than five times the number determined by the driving distance, consider using alternative networks with a lower latency to link the two sites. For example, consider two sites for which the actual driving distance is 135 miles (200 km). Using the routed network calculation, this distance equates to 450 km, or a one-way delay of 2.25 milliseconds. If the network ping test reports latency that is 10 times greater (22.5 ms), then consider other network options that have a lower latency.

Oversubscription can increase intersite network latency. To check for oversubscription, look for the packet loss ratio. The packet loss ratio indicates the need to retransmit data across the intersite link. Each retransmission delays the data in the queue behind the current packet, which increases the completion time for the pending transmissions. HP Continuous Access EVA does not support a packet loss ratio greater than 0.2% after the addition of replication traffic.

Using the network latency results, determine if the I/O per second (IOPS) and throughput will meet your application requirements. If performance does not meet your requirements, you must reconfigure the applications and/or the intersite link between sites (for example, put applications on different links, change the intersite link technology, or adjust the distance). For information about calculating IOPS and throughput performance, see the *HP StorageWorks Continuous Access EVA performance estimator user guide*, which is available at the following web site:

http://h20000.www2.hp.com/bizsupport/TechSupport/DocumentIndex.jsp?contentType=SupportManual&locale=en_US&docIndexId=179911&taskId=101&prodTypeId=12169&prodSeriesId=435080

TIP: When adjusting bandwidth or distance, ensure you are using the peak, not the average, requirements of the application.

Balancing data replication mode and performance

This section defines replication and its effect on performance. Use this information to determine the type of replication required for your environment. There are two replication modes–asynchronous and synchronous.

Topics include:

Replication overview

Replication occurs when a host sends write I/O to the source disk. The controller intercepts the write and sends a copy to the destination disk. As part of the replication process, the EVA controller software adds a message with a unique sequence number to the replication I/O. This number is unique within a DR group. The destination controller is responsible for verifying the sequence number and requesting any missing data before applying the current write. On initialization of the connection between the controller and the switch, the controller configures the switch for in-order delivery of all messages. If a problem occurs, such as a single message not being delivered, the controller of the source disk resends the missing message to ensure in-order delivery of write I/O.

The controller software also:

- Attempts to balances performance between replication and non-replication disks so that replication disks do not occupy all of the controller resources.
- Examines new writes received while a full copy is in progress and ensures the data is not replicated if the block range of a write is within the area that has yet to be copied.

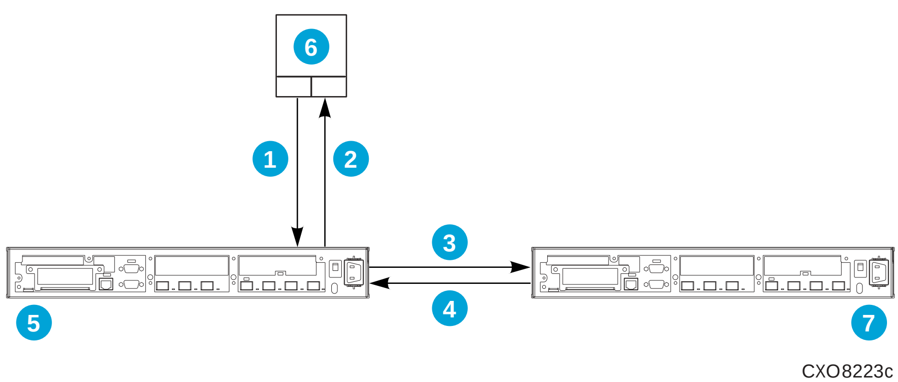
There are two replication modes: asynchronous and synchronous. You set the replication mode when you create a data replication (DR) group.

Asynchronous replication

Asynchronous replication occurs as follows (Figure 3):

1. The server receives data and stores it in the source controller cache.
2. The source controller returns an I/O completion acknowledgement to the server.
3. The source controller sends the data to the destination controller cache.
4. When the destination controller receives and stores the data in cache, it returns an acknowledgement to the source controller.

Typically, asynchronous replication is chosen when response time is a priority.



1	Write I/O	2	Acknowledgement
3	Write I/O	4	Acknowledgement
5	Source controller	6	Host
7	Destination controller		

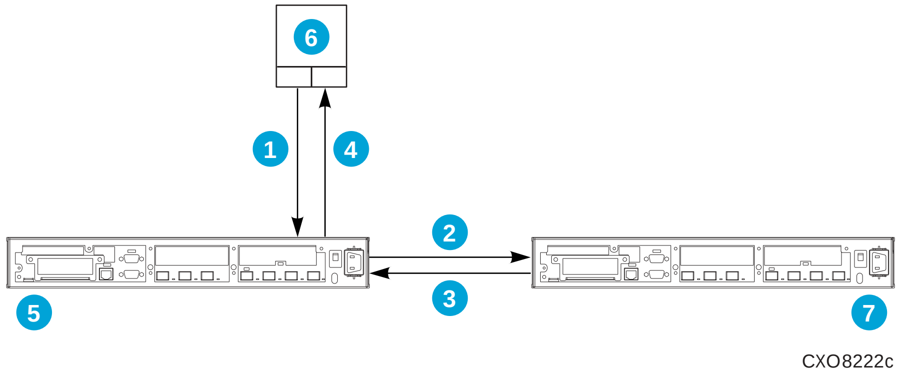
Figure 3 Asynchronous replication sequence

Synchronous replication

Synchronous replication occurs as follows (Figure 4):

1. The server receives data and stores it in the source controller cache.
2. The source controller sends the data to the destination controller cache.
3. When the destination controller receives and stores the data, it sends an acknowledgement to the source controller.
4. When the source controller receives the acknowledgement from the destination controller, it returns an I/O completion acknowledgement to the server.

Typically, synchronous replication is chosen when data currency is a priority.



1 Write I/O
2 Write I/O
3 Acknowledgement
4 Acknowledgement
5 Source controller
6 Host
7 Destination controller

Figure 4 Synchronous replication sequence

Recommendation

HP recommends the use of synchronous replication, whenever possible, for the following reasons:

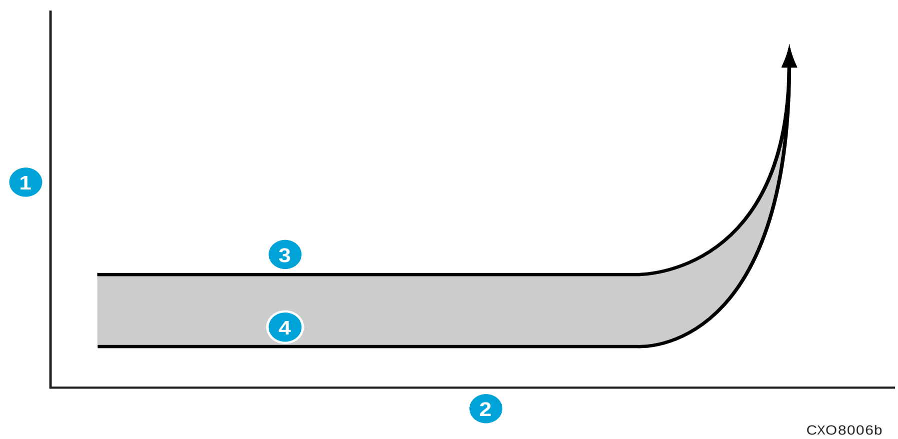
- Whether the replication is performed before the write is acknowledged (synchronous), or after the write is acknowledged (asynchronous), both modes use the same buffers to move the data to the destination array. Therefore, asynchronous replication only improves response time, not performance.
- Both replication modes are queued at the source array. The queue size is finite and the peak rate at which the queue can be emptied is limited by the separation distance, not the type of replication. Both synchronous and asynchronous replication support the same maximum separation distance.

TIP: Typically, the limiting factor in replication performance is the distance between sites, not the bandwidth of the communications link.

- Synchronous replication provides data protection and asynchronous mode offers no additional throughput or performance capabilities over synchronous mode. This differs from other solutions in which asynchronous replication is required when the separation distance is past a typical metropolitan threat separation.
- Although asynchronous replication can reduce the response time (write completion back to the host), it puts each of the outstanding writes at risk if the source array is lost between the time the write is acknowledged to the host and it is written at the destination.
- Asynchronous replication does not use the write history log for its buffer. A new asynchronous replication request is temporarily changed to a synchronous request when the replication buffer is full.

Figure 5 illustrates the performance differences between synchronous and asynchronous replication. The vertical axis is the response time (the time it takes to complete a single I/O over distance for both types of replication). The horizontal axis is the relative I/O rate for a given separation distance. Only when the application I/O rate is below saturation (the shaded area) will asynchronous replication respond faster than synchronous replication.

NOTE: Because actual values depend on the size of the I/O and the intersite distance, only the relative relationship between response time and I/O rate is shown in Figure 5.



1	Response time	**2**	I/O rate
3	Synchronous	**4**	Asynchronous

Figure 5 I/O rate saturation

Optimizing resources

This section describes ways you can optimize your resources for replication. Topics include:

- Bidirectional replication, page 21
- Multiple replication relationships, page 22

Bidirectional replication

Bidirectional replication means that an array can have both source and destination disks, but the disks must belong to separate or unique DR groups. A single virtual disk cannot be both a source and destination. For example, you can configure one DR group to replicate data from array A to array B, and another DR group to replicate data from array B to array A (Figure 6). This configuration does not affect normal operations or your failover policy. Further, bidirectional replication enables you to actively use the destination array while providing a disaster-tolerant copy of the source array's data.

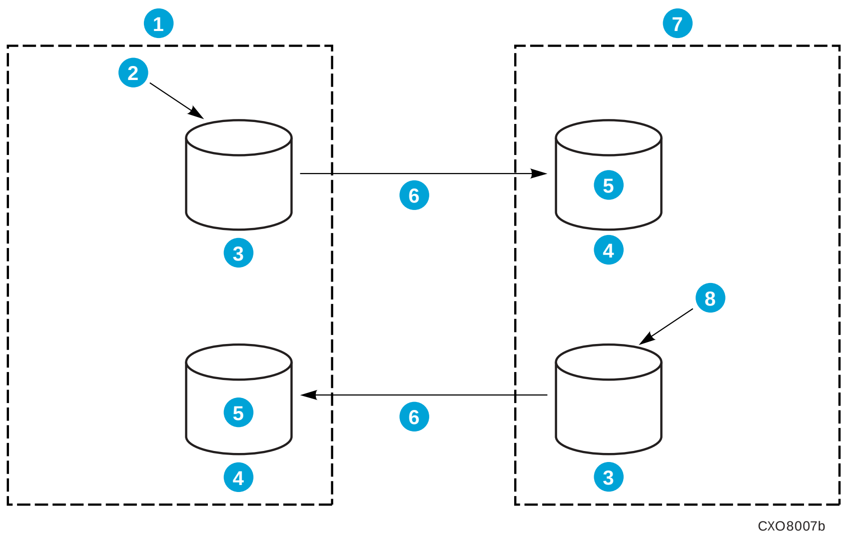
TIP: When performing bidirectional replication, the disk groups on both arrays should be the same size and type.

If your business needs require bidirectional data transfers, determine the effect it will have on the intersite links. For example, consider the bandwidth as two unidirectional flows. Then, use the sum of these two flows as the maximum amount of bandwidth, and then add the two flows together to obtain the worst-case

bandwidth requirements in either direction. The worst-case scenario occurs during recovery after failover and should be used in delivering intersite bandwidth requirements.

You can configure bidirectional replication so that hosts at the destination site, while performing secondary tasks, are ready to support the source applications if the local site is damaged. In addition, the remote site servers can be configured to handle some of the local site application load, if the application load is easily divided. Secondary tasks that can be performed by the remote site include backup, report generation, and data mining.

Figure 6 illustrates bidirectional replication.



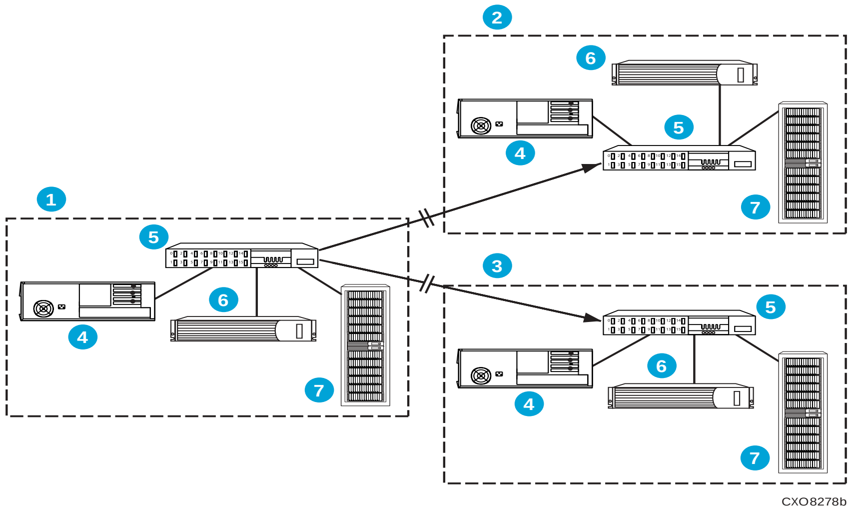
1 Array A
2 Application 1 writes
3 Source
4 Destination
5 Copy
6 Transfer direction
7 Array B
8 Application 2 writes

Figure 6 Bidirectional replication

Multiple replication relationships

One array can have a replication relationship with a maximum of two other arrays (Figure 7). There are two unique sets of data (one set of data for each relationship) and not three copies of the same data. For example, site A replicates to site B (one unique set of data) and site A replicates to site C (another unique set of data). Using either synchronous or asynchronous replication provides similar levels of performance.

Evidence shows that multiple replication relationships can improve performance when the overall performance is limited by distance, and provided there is sufficient link bandwidth and capability in the array, FCA, and servers. For more information about performance and other relationship examples, see the *HP StorageWorks Continuous Access EVA performance estimation user guide*.

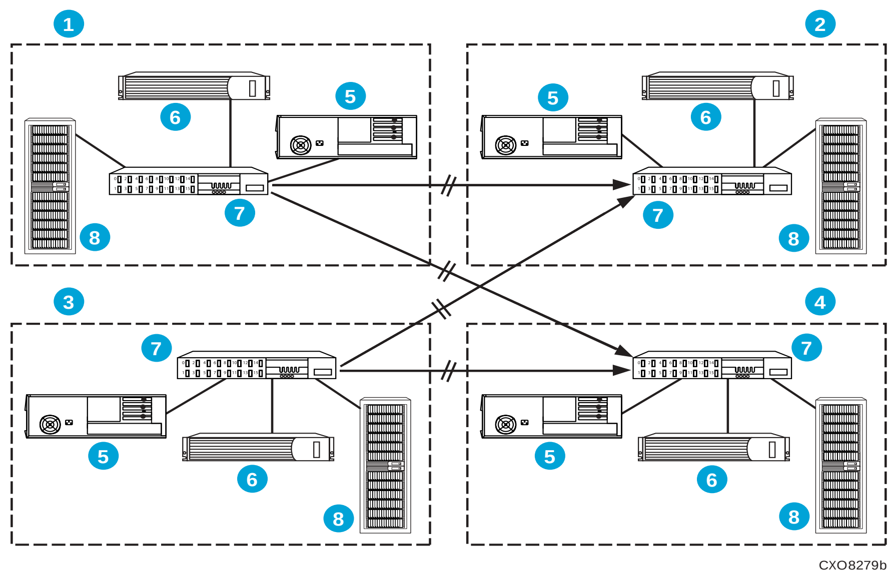


1	Site A	**2**	Site B
3	Site C	**4**	Hosts (one at each site)
5	Management servers (one at each site)	**6**	Switches (one at each site)
7	Arrays (one at each site)		

Figure 7 One source replicating to two destinations

It may be possible to increase performance across a link when using four arrays, two at each site. In this configuration, the pair of source arrays share the two destination arrays. The sharing of resources improves the performance over that of two single replication relationship pairs, each having a dedicated destination. See Figure 8 and Figure 9 for comparison.

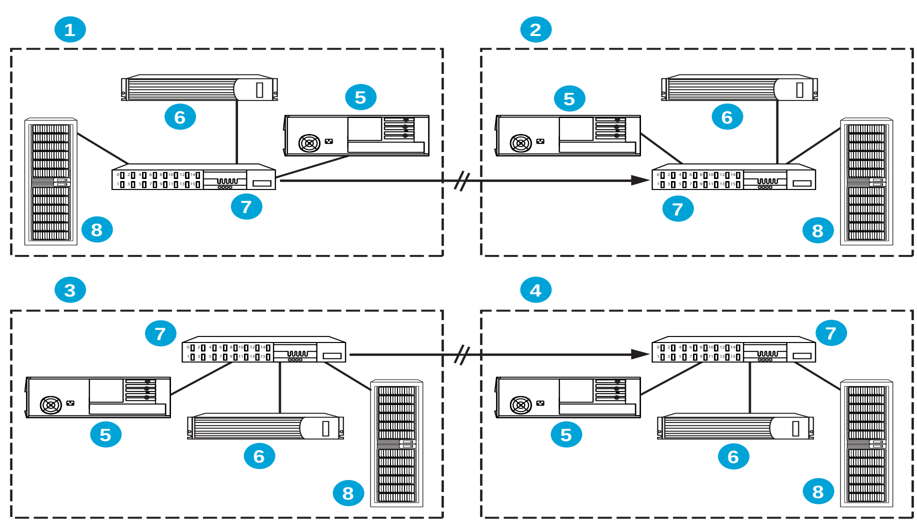
NOTE: While this solution optimizes performance, it requires careful planning as each link supports different DR groups.



1 Site A
2 Site B
3 Site D
4 Site C
5 Hosts (one at each site)
6 Management servers (one at each site)
7 Switches (one at each site)
8 Arrays (one at each site)

Figure 8 Four relationships sharing destinations

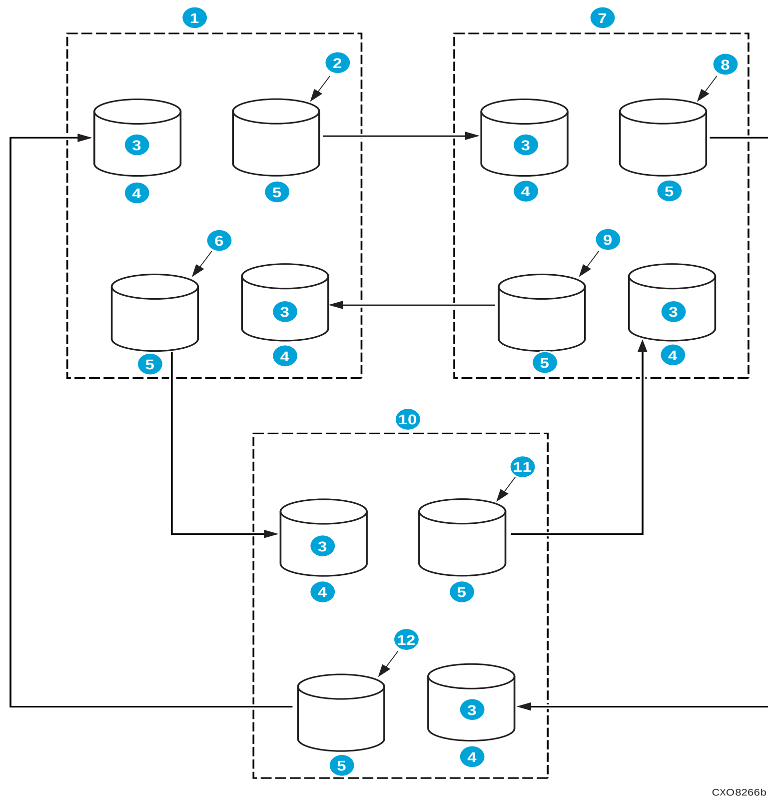
In Figure 9, sites A and D have dedicated destinations—sites B and C, respectively.



1 Site A
2 Site B
3 Site D
4 Site C
5 Hosts (one at each site)
6 Management servers (one at each site)
7 Switches (one at each site)
8 Arrays (one at each site)

Figure 9 Two relationships with dedicated destinations

Although HP Continuous Access EVA supports multiple relationship configurations (Figure 10), HP does not test or recommend these types of configurations.



1	Array A	2	Application 1 writes
3	Copy	4	Destination
5	Source	6	Application 6 writes
7	Array B	8	Application 4 writes
9	Application 2 writes	10	Array C
11	Application 3 writes	12	Application 5 writes

Figure 10 Multiple replication relationship not recommended by HP

3 Planning the array configuration

This chapter describes how to plan the array configuration for remote replication. Topics include:

- Data replication groups, page 27
- Performance considerations, page 27

Data replication groups

A data replication (DR) group is a named group of virtual disks selected from one or more disk groups so that they remotely replicate to the same destination, fail over together, share a log (DR group log), and preserver write order within the group. DR groups are created in pairs, consisting of a source DR group and a destination DR group. Each DR group may contain one or more virtual disks. For optimal performance, however, limit each DR group to one virtual disk. One DR group supports a maximum of eight source-destination pairs from one or more disk groups within the same array.

Continuous Access EVA supports a maximum of 128 DR groups on each array, with each DR group consisting of one virtual disk. If you create DR groups with multiple virtual disks, the maximum decreases. For example, if each DR group contains eight virtual disks, the maximum number of DR groups decreases to 16.

It is not necessary to use all 128 DR groups, nor will smaller HP Continuous Access EVA configurations need all of them. For example, one application can generate multiple I/O streams using the full bandwidth of the controller and only require five DR groups. Further, the size of a single virtual disk can require most of the available capacity within the storage array cabinet. DR groups can exist in any supported Vraid0, Vraid1, or Vraid5 format, but the source and destination in each DR group must be of the same type.

NOTE: In a Continuous Access EVA configuration, HP recommends that you do not use Vraid 0 at any time and only use Vraid 5 on arrays with fewer than eight disks.

Snapshots and snapclones share some of the virtual disk's resources that are also used for replication. Therefore, HP recommends that you limit the number of snapclones created when peak replication performance is required. With Virtual Controller Software (VCS) version 3.02, you can choose the Vraid type for both the snapclone and the snapshot, as well as the location of the snapclone. One DR group supports a maximum of eight snapshots or snapclones.

Performance considerations

The section describes the performance issues to consider as you plan disk group configuration for replication. Topics include:

- Performance requirements and I/O patterns
- Write history logs

For instructions about configuring disk groups, such as creating disk groups and designing capacity, see the *HP StorageWorks Command View EVA user guide*, which is available at the following web site: http://h18006.www1.hp.com/products/storage/software/som/index.html.

Performance requirements and I/O patterns

In HP Continuous Access EVA, performance is not always related to the number of disks available to store data. Because of replication to a remote array, there is a point where adding more disks does not increase the maximum write rate, only the maximum random read rate. If an application has a high percentage of random reads compared to writes, then a large number of disks in the disk group is appropriate. If, however, there is a high percentage of writes compared to reads, the actual task of replication limits performance rather than a limited number of disks. Additionally, sequential access (read or write) is limited by the per-disk performance rather than the number of disks in the disk group.

Analyze the transfer size of the replication I/O and the distance between the source and destination arrays. Based on the results, consider the following performance issues:

- If the application I/O stream is dominated by a mix of simultaneous sequential and random transfers, determine how these streams can be directed to specific virtual disks.
- Put virtual disks with similar tasks in the same disk group. In general, separate sequential I/O stream data (database logs, rich content) from random I/O streams (database information store, file shares).
- Note that transfer profiles to a given disk that differ over time are not a major consideration. A virtual disk that receives sequential transfers for part of the day and random accesses for the rest of the day operates well in both cases. The issue is accommodating simultaneous sequential and random streams.

Write-to-disk rate

For current applications, determine the I/O performance rates of each virtual disk replicated to the remote site. Calculating the average and maximum write rate (write I/O per second) and peak write transfer rate (bytes per second) can require some analysis and time. Use operating system-specific tools, such as *PERFMON* or *IOSTAT*, to collect the data. If the data is available only as an average over time, attempt to ascertain the peak hours of operation and estimate the peak write rate and write transfer rates.

Record the peak and average write rates and the write transfer rates. If bidirectional, record these numbers for each direction. Compare these numbers with intersite link technologies to determine which technology is most cost-effective. If there is not one technology that proves most cost-effective, consider other methods of replicating data or replicate only the most critical data, such as transaction or retransmission logs.

The average load on any link must not exceed 40% of rated capacity, and the peak loading must not exceed 45% of rated capacity. This limitation allows I/O from a failed link or fabric to run on the active fabric or link, without causing additional failures of the surviving fabric.

One final consideration is that as distance (or intersite latency) increases, the distance limits the maximum I/O rate more than the bandwidth. For example, if the calculations used in the previous examples were repeated using 100 msec of one-way latency instead of 1msec, regardless of the bandwidth, the math shows only five writes per second are possible for a single replication stream.

Write history logs

A DR group uses the write history log when a problem occurs with the intersite links. The write history log requires additional space, ranging from 136 MB to 2 TB. The controller places the log in the disk group with the most free space. This may be the disk group containing the DR group, or a different group. In either case, the space must be included in the disk space planning. Logs are of type Vraid1.

VCS 3.02 and later allows you to create disk groups using near-online and online disks. Near-online disks are more cost effective for storing the write history logs for all DR groups that exist on an array. VCS 3.02 and later automatically selects a near-online-based disk group, if one exists when the DR group is created, for the write history log. The write history log will not move if additional disk groups are created after the DR group is created.

Table 4 lists cases that illustrate the process VCS uses to select a disk group for the write history log.

Table 4 Disk group selection process

Case	Action
The array contains one defined disk group.	VCS automatically places the write history log in the defined disk group.
The array contains one near-online disk group and more than one online disk group.	VCS automatically places the write history log in the near-online disk group.
The array contains multiple near-online disk groups.	VCS selects the near-online disk group containing the most free space. If more than one near-online disk group has the same amount of free space, VCS alternates between the members to balance the use of the free space.
The array contains multiple online disk groups and no near-online disk groups.	VCS selects the online disk group containing the most free space. If more than one online disk group has the same amount of free space, VCS alternates between the members to balance the use of free space.

4 Planning the fabric configuration

This chapter describes all supported Continuous Access EVA fabric configurations. Topics include:

Basic Continuous Access EVA over fiber

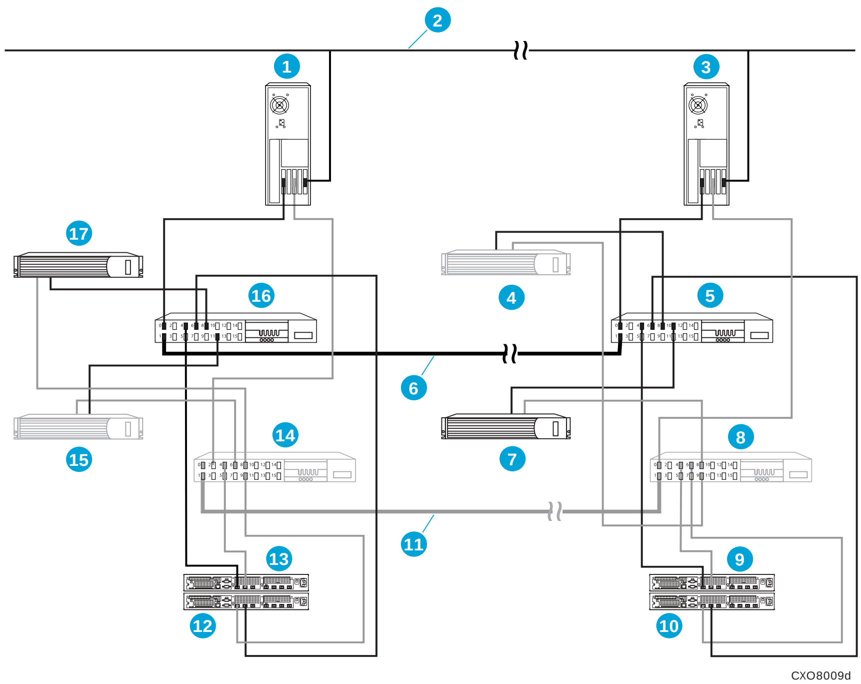
Continuous Access EVA over fiber is the most basic HP Continuous Access EVA configurations. As shown in Figure 11 on page 32, this configuration supports two redundant fabrics, in which the first FCA in host A is connected to switch W, and the second FCA in host A is connected to switch X. The top controller of the array on the left is attached to switch W and switch X, and the bottom controller is also attached to switch W and switch X.

At the right side of the figure, the backup host and the array are wired the same way as at the left side. Using the same switch ports for the same functions at both sites reduces confusion during a disaster or debugging. HP also recommends naming the two fabrics to distinguish them. For example, name them top and bottom or black and gray.

This dual-fabric SAN provides no single point of failure at the fabric level. For example, broken cables, switch updates, or an error in switch zoning can cause one fabric to fail, leaving the other to temporarily carry the entire workload. Currently, a maximum of 28 B-series switches, 16 C-series switches, or 24 M-series switches are supported in each fabric. For additional information, contact your local HP representative. Non-Continuous Access EVA servers and storage are allowed on each fabric, if each is kept in a zone separate from the Continuous Access EVA solution space.

Total solutions larger than what is possible with a single supported solution are also supported. Each of the smaller solution instances must exist within a single management zone that conforms to all the requirements outlined in the section "Configuration rules" on page 34. The combination of two or more

solution instances must not exceed the maximum configuration described in the *HP StorageWorks SAN design reference guide*.



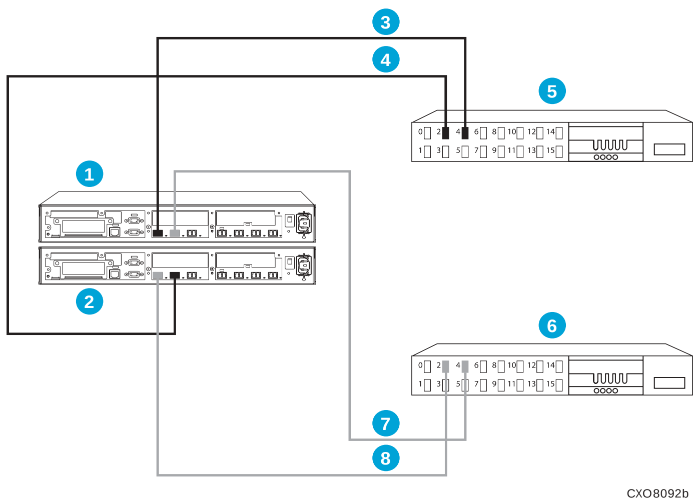
1 Host A	**2** Network interconnect
3 Host B	**4** Management server
5 Switch Y	**6** ISL - black fabric
7 Management server	**8** Switch Z
9 Controller B1	**10** Controller B2
11 ISL - gray fabric	**12** Controller A2
13 Controller A1	**14** Switch X
15 Management server	**16** Switch W
17 Management server	

Figure 11 Basic Continuous Access over fiber configuration

Cabling

Figure 12 on page 33 shows the supported cabling. The basic rule is that the first or left-hand port on the top controller is cabled to the first fabric, and the other port of the same controller is cabled to the other fabric. The other (bottom) controller is cabled so that the left-hand port is attached to the second fabric, while the second port is cabled to the first fabric; the opposite of the first (top) controller. Even though it does not matter which switch ports are used, symmetry is recommended. If there is a fabric failure, the

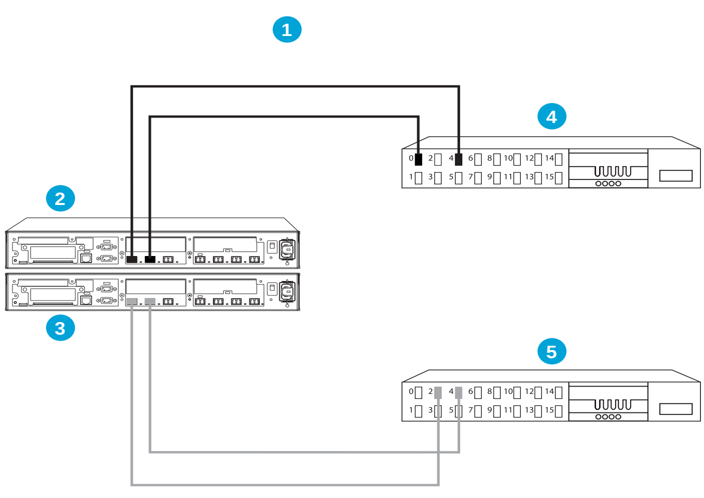
virtual disk stays on the same controller but moves to the other fabric. Any other controller-to-fabric cabling scheme is not supported.



1	Controller A	**2**	Controller B
3	A(1)	**4**	B(2)
5	Switch 1	**6**	Switch 2
7	B(1)	**8**	A(2)

Figure 12 Supported cabling

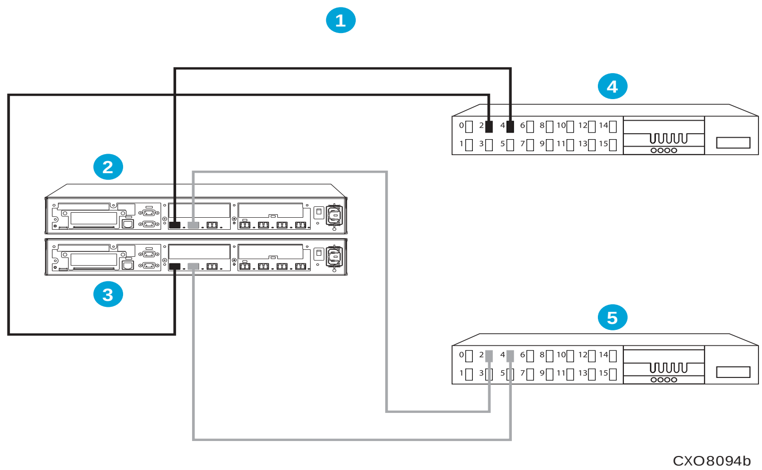
The cabling in Figure 13 is not supported because both ports of a controller are on the same fabric, and the virtual disk changes controllers if a fabric failure occurs.



1	This cabling is not supported	**2**	Controller A
3	Controller B	**4**	Switch 1
5	Switch 2		

Figure 13 Example 1 of unsupported cabling

The cabling in Figure 14 is not supported because both port 1s and port 2s share the same fabric. This inhibits static load balancing by DR group and limits failover operations.



1 This cabling is not supported
2 Controller A
3 Controller B
4 Switch 1
5 Switch 2

Figure 14 Example 2 of unsupported cabling

Configuration rules

The following rules apply to the basic Continuous Access EVA over fiber configuration:

- At least two, but no more than 16 arrays can be split between the local and remote sites. Each array dedicated to an HP Continuous Access EVA configuration must have dual array controllers.
- The operating system on each host must either implement multipath support (such as OpenVMS and Tru64 UNIX) or support it using HP StorageWorks Secure Path software.
- The minimum HP StorageWorks Enterprise Virtual Array storage configuration supports a maximum of 28 disks on each array, with larger configurations supporting up to 240 disks (with an expansion cabinet). Destination virtual disks must have the same geometry and capacity as the source virtual disks of the source-destination pair.
- Source and destination disk groups do not require the same geometry (online and near-online) and capacity. However, the disk group with fewer disks has a lower peak performance than the one with more disks. Disk groups supporting bidirectional replication should be symmetric in both size and disk performance for consistent performance from either array.

NOTE: A disk group should contain only one model of physical disk.

- Both controllers within an array must have the same version of the Virtual Controller Software (VCS) installed and running. The exception to this rule is that when changing the firmware, both arrays must have the same version of VCS installed and running.
- A minimum of two FCAs (or one dual-port FCA) are required for each host to ensure that no single point of failure exists between the host and the array. A maximum of four single-port, two single-port and one dual-port, or two dual-port FCAs per host is allowed.
- A maximum of 256 FCAs per array are allowed. The ports can be a mix of single- and dual-port FCAs. At two FCA ports per server, the 256 limit equates to a maximum of 128 servers.

- To maintain write order across the members of a DR group and to maintain a fail one/fail all model, all members of the DR group must be preferred to the same array controller and use the same FCA port pair. In a configuration with multiple pairs of FCAs, all the virtual disks belonging to the same DR group are restricted to using only one FCA port per host.
- Each site should have one HP Continuous Access EVA documentation set and one array platform kit (appropriate for the array type and server operating system) per implemented operating system platform. The reason is to support disaster recovery, rebuilding or repair of the surviving system should access to the other system or site not be possible.
- One GBIC or SFP that is appropriate for the type of fiber optic cable being used is required per switch port connection.
- Each site must have at least one management server. Two management servers are recommended for high availability.
- The designed switch configuration must be supported. See the *HP StorageWorks SAN design reference guide*.

NOTE: Use a third-party vendor to acquire and install all SMF optic cables, any MMF optic cables longer than 50 m, and WDM interfaces.

Maximum Continuous Access EVA over fiber

The maximum Continuous Access EVA over fiber configuration supports up to 16 arrays split between the two sites, depending on the SAN topology. High performance SANs can be built using a *skinny tree* topology as defined in the *HP StorageWorks SAN design reference guide*. However, using a high-performance SAN can reduce the maximum number of hosts and arrays, due to the reduction in open switch ports.

A large port-count SAN can be constructed using multiple instances of smaller configurations, each in a separate management zone, subject to the limits in fabric sizes as defined in the *HP StorageWorks SAN design reference guide*.

NOTE: See Table 6 on page 51 for limits based on separation distance.

Continuous Access EVA over fiber with long-distance GBICs and SFPs

Basic Continuous Access EVA over multimode fiber supports distances of up to 500 m at 1 Gb/s, and up to 300 m at 2 Gbps.

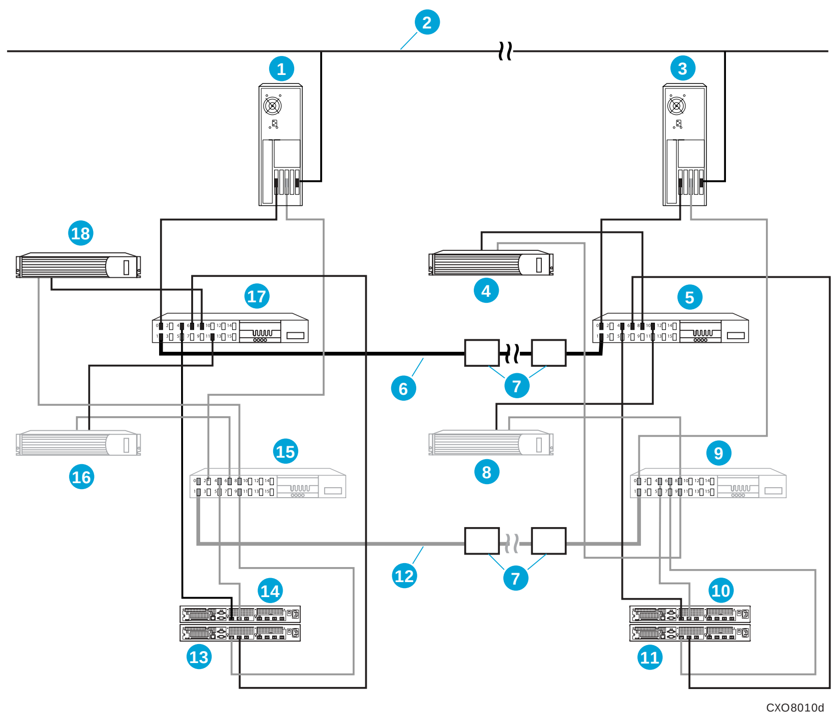
Longer distances require the use of long-distance and very long-distance GBICs and SFPs. Long-distance GBICs and SFPs using single-mode 9-µm fiber can span distances up to 10 km. At the time of this publication, very long-distance GBICs and long-distance SPFs running on 9-µm fiber can support distances up to 100 km at 1 Gbps, or up to 35 km at 2 Gbps.

B-series switches may require the optional Extended Fabric License for this configuration.

Continuous Access EVA over WDM

As an option, Continuous Access EVA over fiber also supports the use of wavelength division multiplexing (WDM) instead of the long-wave or very long-distance GBICs.

Figure 15 shows a Continuous Access EVA over WDM configuration. Currently, HP supports HP Continuous Access EVA over any vendor's dense wavelength division multiplexing (DWDM) or coarse wavelength division multiplexing (CWDM) system, if the installation conforms to vendor specifications. There will be a performance impact due to distance and / or limited buffer-to-buffer credits. In addition the maximum distance may be limited by the switch vendor. See the *HP StorageWorks Continuous Access and Data Replication Manager SAN extensions reference guide* for more information.



1	Host A	**2**	Network interconnect
3	Host B	**4**	Management server
5	Switch Y	**6**	ISL - black fabric
7	WDM	**8**	Management server
9	Switch Z	**10**	Controller B1
11	Controller B2	**12**	ISL - gray fabric
13	Controller A2	**14**	Controller A1
15	Switch X	**16**	Management server
17	Switch W	**18**	Management server

Figure 15 Continuous Access EVA over WDM configuration

The difference between the use of WDM and the basic solution is the replacement of at least one, if not both, long-distance GBICs and single-mode fiber with a multiplex unit, shortwave GBICs, and multimode fiber. For more information on Continuous Access over WDM, see the *HP StorageWorks Continuous Access and Data Replication Manager SAN extensions reference guide*.

Configuration rules

The following rules apply to the HP Continuous Access EVA over WDM configuration:

- Typically, one switch-to-WDM interface cable is required per wavelength of multimode fiber to connect the switch to the WDM unit.
- If you are using older B-series switches, an Extended Fabric License may be recommended. See the *HP StorageWorks Continuous Access and Data Replication Manager SAN extensions reference guide* for more information.

Extended Continuous Access EVA over IP configuration (long-distance solution)

The extended Continuous Access EVA over IP configuration is similar to the basic Continuous Access EVA configuration except for the use of Fibre Channel-to-IP gateways. Due to the dual fabrics, two gateways are required at each site—one per fabric, for a total of four per solution, dedicated to that solution and eliminates single points of failure.

Continuous Access EVA over IP has the same maximum configuration limits as those described in "Configuration rules" on page 34. Multiple instances can share the same fabric if all components are in unique management zones and the network bandwidth is sufficient for all traffic flowing between the sites in a worst-case scenario.

NOTE: See Table 6 on page 51 for limits based on separation distance. Also, see the *HP StorageWorks Continuous Access and Data Replication Manager SAN extensions reference guide* for information about network requirement differences for single ISL and shared or dual ISLs.

Current versions of the Fibre Channel-to-IP gateways support direct connection to either 10/100-Mbps copper or 1-Gbps optical Ethernet. The Fibre Channel-to-IP (FCIP) gateway uses the intersite network bandwidth that is set aside for the storage interconnect. The IP tunnel that is created to support the FCIP traffic also provides enhanced security of the data because of the nature of IP tunnels. HP recommends designing the IP tunnels with enough bandwidth to carry all the intersite data traffic in case either link fails.

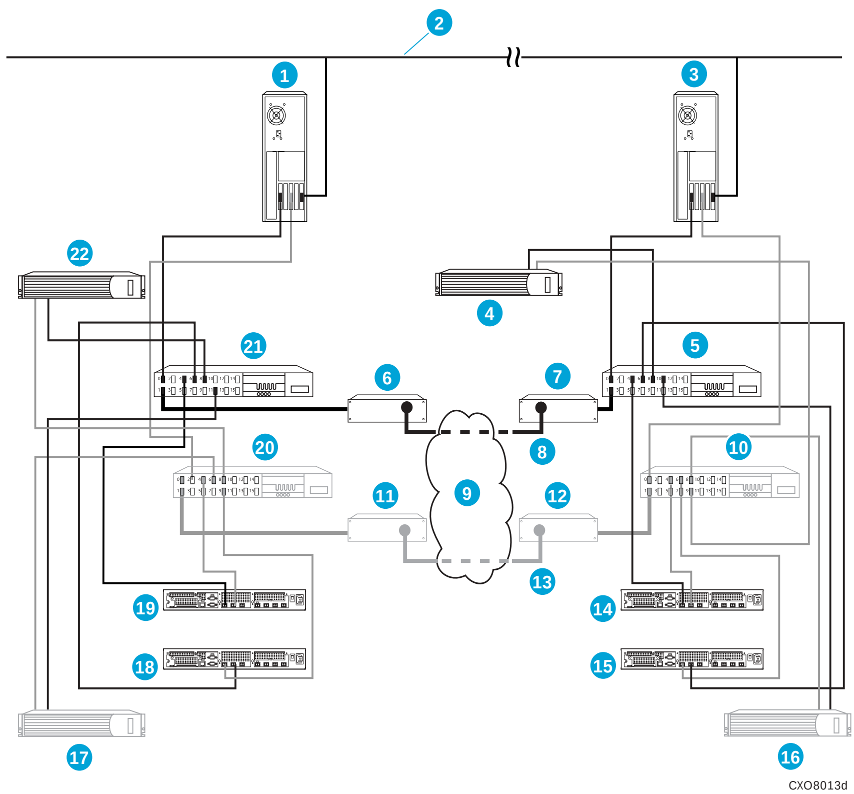
Packet loss within the IP network is tolerated and even expected, due to oversubscription or inadvertent congestion. As a result of these errors, a packet can be dropped without the sender being notified that the packet is now lost. However, for every packet that is lost in the intersite network, that same packet must be retransmitted by the sender until received at the receiver or until a link timeout occurs. The effect of these retransmissions is seen as increased delay on the particular packet and a resulting decrease in the available bandwidth because it is not available for new packets. The greater the percentage of packets lost in the transfer, the lower the effective bandwidth and the longer a particular transfer will take. For example, using a maximum bit error rate (BER) of 1 in 10^{10}, one in approximately 500,000 2-KB data frames will require retransmission. At the other extreme, a 1 percent packet loss translates to losing 1.3 data frames in 100 2-KB packets, or some part of almost every 64-KB write due to network errors.

By way of comparison, Fibre Channel networks are designed for a BER of 1 in 10^{12}, or one in approximately 50,000,000 2-KB data frames. Because of this, HP recommends that both a maximum delay and maximum BER be specified in the network service contract. See the *HP StorageWorks Continuous Access and Data Replication Manager SAN extensions reference guide* for current network delay and packet loss ratio requirements.

Some wide area networks can be built using a ring architecture where one way around the ring is significantly shorter in time than the other (longer) way around the ring. Other wide area networks can also have two paths of different lengths, a shorter and a longer one. In either case, Continuous Access EVA supports these occasional drastic changes in the intersite delay, if the longest delay does not exceed

an average of 100 ms. To accomplish this, the EVA storage system firmware periodically tests the intersite delay and adjusts the heartbeat rates, message time-outs, and outstanding I/O counts for optimum performance of an intersite link, based on the current intersite delay. The latency jitter specification in the *HP StorageWorks Continuous Access and Data Replication Manager SAN extensions reference guide* refers to either of the two options in the ring and not the difference in latency of one segment of the ring versus the other.

Figure 16 shows a Continuous Access EVA over IP configuration.



1	Host A	**2**	Network interconnect
3	Host B	**4**	Management server
5	Switch Y	**6**	FC-IP A
7	FC-IP Y	**8**	ISL - black fabric
9	IP	**10**	Switch Z
11	FC-IP B	**12**	FC-IP Z
13	ISL - gray fabric	**14**	Controller B1
15	Controller B2	**16**	Management server
17	Management server	**18**	Controller A2
19	Controller A1	**20**	Switch X
21	Switch W	**22**	Management server

Figure 16 Continuous Access EVA over IP configuration

Additional configuration rules

Consider the following requirements when designing a Continuous Access EVA over IP configuration:

- Typically, one multimode fiber is required to connect the switch to the FCIP gateway.
- Some FCIP gateways are supported only on the older B-series switches and require the Remote Switch Key (vendor-dependent).

A third-party vendor is used to acquire and install all SMF optic cables, any MMF optic cables longer than 50 m, and the FCIP interface boxes.

Some IP gateways provide a mechanism to notify the fabric that connectivity to the remote gateway has been lost. Other gateways require the use of a fabric-based heartbeat to detect loss of the intersite IP network connection. Vendors that require the fabric heartbeat require installation of the Remote Switch Key license onto those two switches that directly connect to the IP gateway. See the *HP StorageWorks Continuous Access and Data Replication Manager extensions reference guide* for more information.

NOTE: The Remote Switch Key is available only on the B-series switches. For those gateways requiring the Remote Switch Key, and on those switches where the Remote Switch Key is installed, do not enable suppression of F-Class frames. Doing so limits the supported size of the Continuous Access EVA over IP SAN to one switch per fabric at each site. See the *HP StorageWorks Continuous Access and Data Replication Manager extensions reference guide* for more information.

NOTE: Regardless of the type of site-to-site transport you use (IP, ATM, SONET), the FC-IP gateway requires either a 10/100 Mb/s copper or a 1 GbE optical interface into the local Ethernet network. The conversion from the local Ethernet to the long-distance network is expected to be performed by a customer-provided network router or gateway.

Continuous Access EVA over SONET

The extended Continuous Access EVA over ATM or SONET configuration is similar to the Continuous Access EVA over IP configuration except for the use of Fibre Channel-to-IP gateways. In Figure 16 on page 38 replace references to Fibre Channel to IP (FC-IP) gateways with FC to SONET gateways.

Continuous Access EVA over ATM

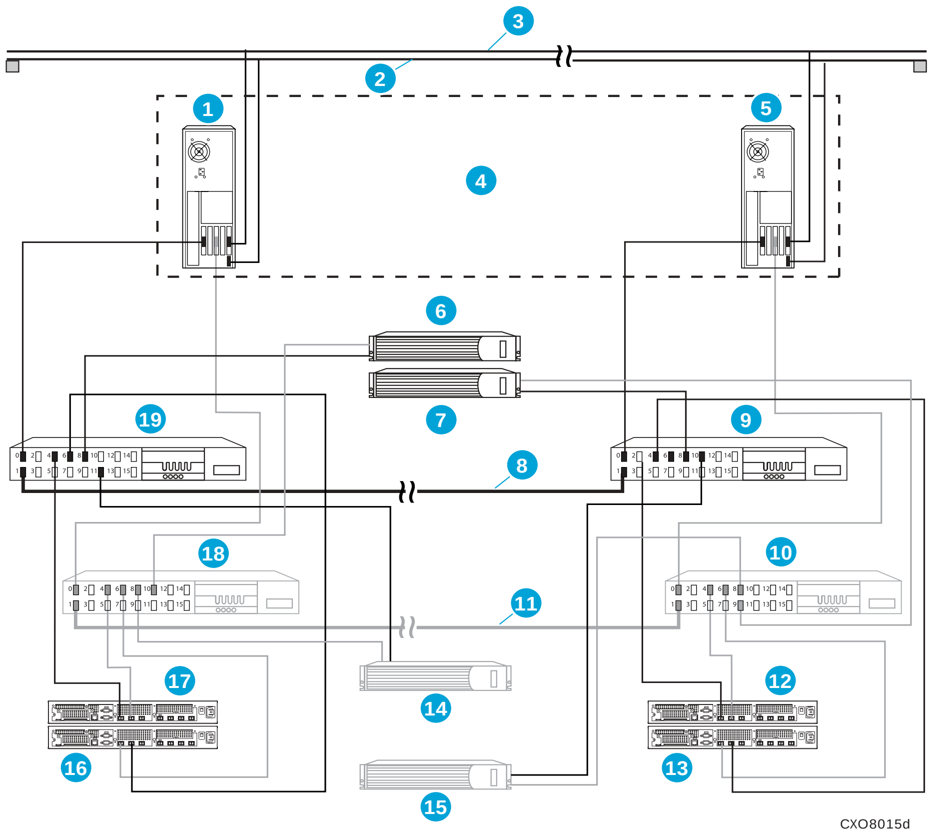
Because no FC to ATM gateways are available at the time of publication, another approach is needed if a solution requires ATM as the transport between the two sites. That approach uses the same FC to IP gateways described in "Extended Continuous Access EVA over IP configuration (long-distance solution)" on page 37, but the IP router is required to provide a blade that interfaces with the ATM network instead of the IP network.

Continuous Access EVA stretched cluster support

Continuous Access EVA supports stretched Microsoft Cluster Servers (MSCS) running Windows 2000, Windows 2003, or Windows NT. In this configuration (Figure 17), half the cluster is at the local site and the other half is at the remote site. If the source host fails, MSCS fails over the application to the surviving host at the remote site and resumes operations using the local site storage.

Applications running in a stretched cluster in server failover mode incur a performance penalty because of the time it takes to read or write data across the intersite link. This performance penalty is directly proportional to the distance between the two sites. During testing, almost no additional effect was observed with separation distances up to 100 km. For more information on stretched cluster support, see the HP ProLiant HA/F500 website at:

http://h18000.www1.hp.com/solutions/enterprise/highavailability/microsoft/haf500/description-eva.html.



1 Host A
2 Cluster heartbeat
3 Network interconnect
4 Multi-site stretch cluster
5 Host B
6 Management server
7 Management server
8 ISL - black fabric
9 Switch Y
10 Switch Z
11 ISL - gray fabric
12 Controller B1
13 Controller B2
14 Management server
15 Management server
16 Controller A2
17 Controller A1
18 Switch X
19 Switch W

Figure 17 Continuous Access EVA stretched cluster configuration

Alternate configurations

The following configurations are supported primarily to reduce the cost of test configurations, and secondarily for production as they do not offer the same level of disaster tolerance and/or high availability as the ”Basic Continuous Access EVA over fiber” on page 31.

Single-fabric configuration

The single-fabric Continuous Access EVA solution is designed for small, entry-level tests or proof-of-concept demonstrations where some distance is needed between each of the two switches in the solution. This solution can also be used for producing copies of data needed for data migration or data mining and for ongoing production where only a single communications link exists.

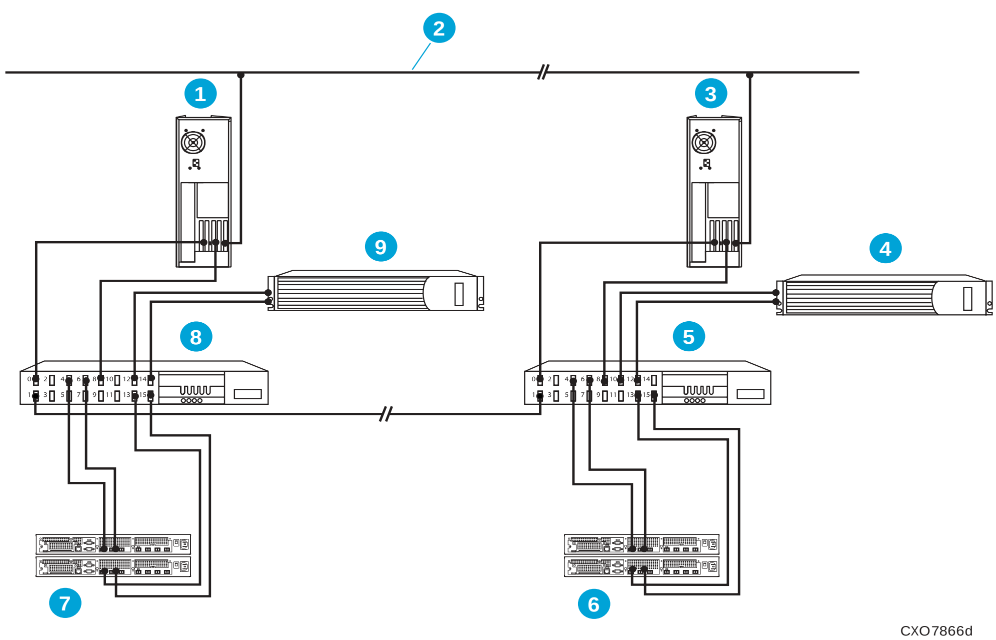
Fabric zoning can be used to create two logical fabrics out of the one physical fabric. Fabric zoning is required to isolate servers as documented in the *HP StorageWorks SAN design reference guide*. These two switches share one intersite link, leaving the remaining ports for hosts, array controllers, and a management server. For example, if a 16-port switch is being used, the remaining 15 ports support up to:

- Four hosts, one array, and one management server
- Two hosts, two arrays, and one management server

An example of the single-fabric configuration using 16-port switches is shown in Figure 18.

Each of the switches shown in Figure 18 can be replaced with any supported fabric topology as defined in the *HP StorageWorks SAN design reference guide*. The same limits apply—up to 28 B-series, 16 C-series, or 24 M-series switches are allowed in the single fabric.

All intersite links supported in the basic Continuous Access EVA are also supported in the single-fabric configuration. This means that the ISL can be direct fiber, a single WDM wavelength, or a single Fibre Channel over IP link.



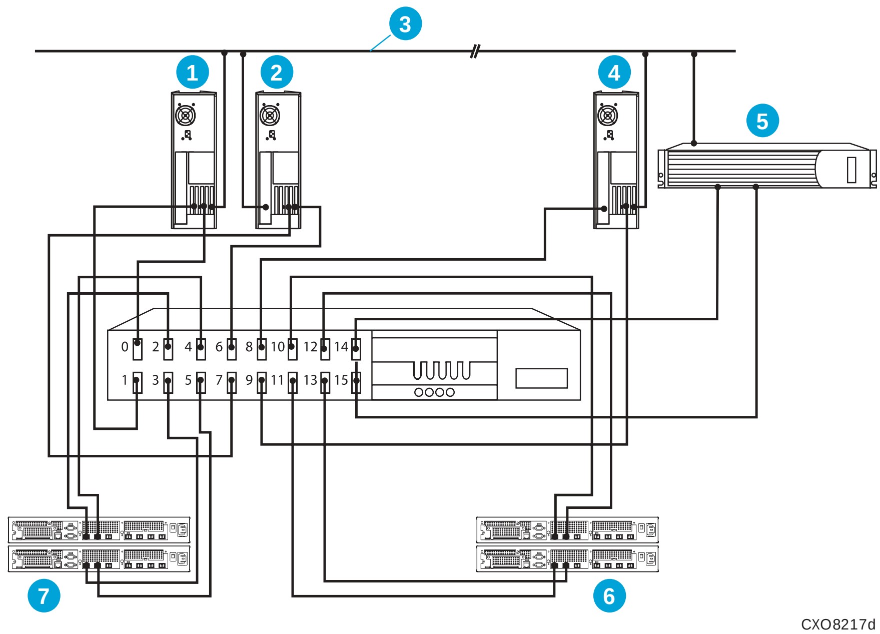
1	Host A	**2**	Network interconnect
3	Host X	**4**	Management server
5	Switch Y	**6**	Controller Y
7	Controller A	**8**	Switch A
9	Management server		

Figure 18 Single-fabric configuration

Single-switch configuration

The single-switch Continuous Access EVA solution is designed for small, single-site, entry-level tests or proof-of-concept demonstrations. This non-disaster-tolerant solution can also be used for producing copies of data needed for data migration or data mining. Dual FCAs and multipathing software are required. A 16-port switch can support a maximum of three hosts, two arrays, and one management server. Large switches support more servers and / or storage arrays if all FCA and array ports are connected to the same switch. Fabric zoning can be used to create the two logical fabrics used by Continuous Access EVA. Fabric

zoning is required to isolate servers as defined in the *HP StorageWorks SAN design reference guide*. An example of the single-switch solution is shown in Figure 19.



1 Host A
2 Host B
3 Network interconnect
4 Host X
5 Management server
6 Controller Y
7 Controller A

Figure 19 Single-switch configuration

In Figure 19, hosts A and B are of one supported operating system and are clustered together using a supported cluster technology for that operating system. In this example, host X is a single server running the same OS as clustered hosts A and B, and therefore is available as a backup to the cluster. As another example, host X with a different OS can be a standalone server used for training on storage failover.

The single switch shown in Figure 19 can be replaced with any supported fabric topology as defined in the *HP StorageWorks SAN design reference guide*. The same limits apply—up to 28 B-series, 16 C-series, or 24 M-series switches are allowed in a single fabric.

Single FCA solution

A host containing a single FCA can be attached to any of the following configurations:

- Basic Continuous Access EVA, and its optional links
- Single fabric
- Single switch

This option allows the use of servers that only support one FCA due to slot restrictions, at the expense of reduced availability due to the single point of failure. To decrease repair time, HP recommends that you deploy some standby hosts, each with a single FCA. If supported, each of these active and standby hosts should be configured to boot from the SAN, so that any standby host can quickly replace any active host.

NOTE: Secure Path is required to mask the redundant path except when using HP Tru64 UNIX or HP OpenVMS.

Advanced configurations

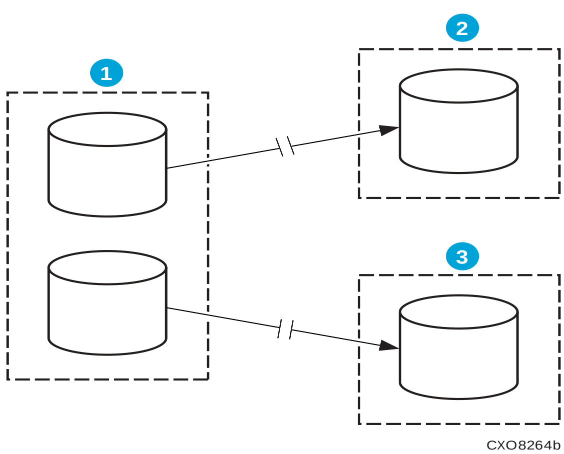
Advanced configurations involve multiple replication relationships, such as:

- Fan-out replication
- Fan-in replication
- Cascaded replication

These relationships are created between the individual arrays. Each relationship supports one or more DR groups not involved in another relationship. Therefore, any one DR group and the source-destination pair within it can only belong to one relationship.

Fan-out replication

In fan-out replication, one DR group is replicated from array A to array B, and another DR group is replicated from array A to array C (Figure 20).

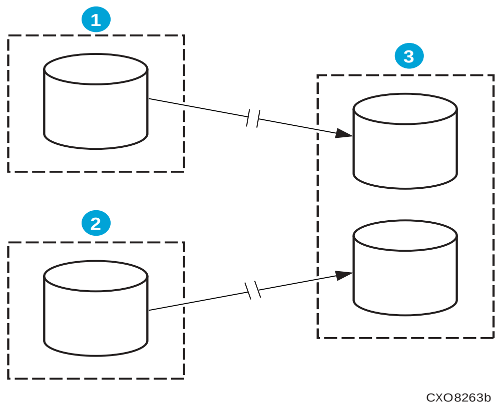


1 Array A
2 Array B
3 Array C

Figure 20 Fan-out replication

Fan-in replication

In fan-in replication, one DR group is replicated from array A to array C, and another DR group is replicated from array B to array C (Figure 21).



1 Array A

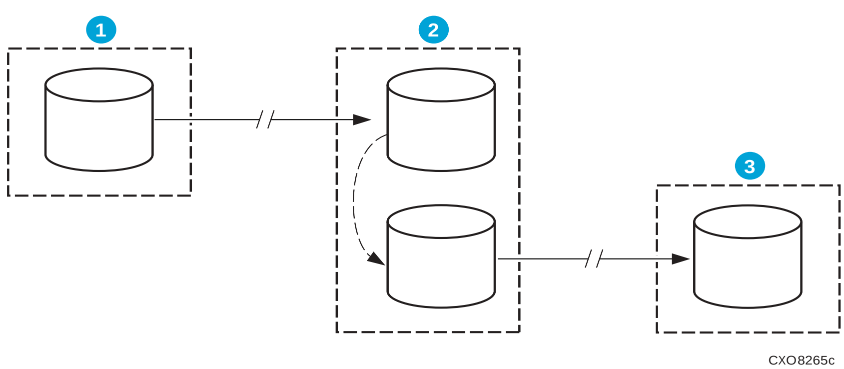
2 Array B

3 Array C

Figure 21 Fan-in replication

Cascaded replication

In cascaded replication, one DR group is replicated from array A to array B, and another DR group is replicated from array B to array C. In this configuration, the source disk for the array B-to-array C replication is a snapclone copy of the destination disk in the array A-to-array B replication (Figure 22).



1 Array A

2 Array B

3 Array C

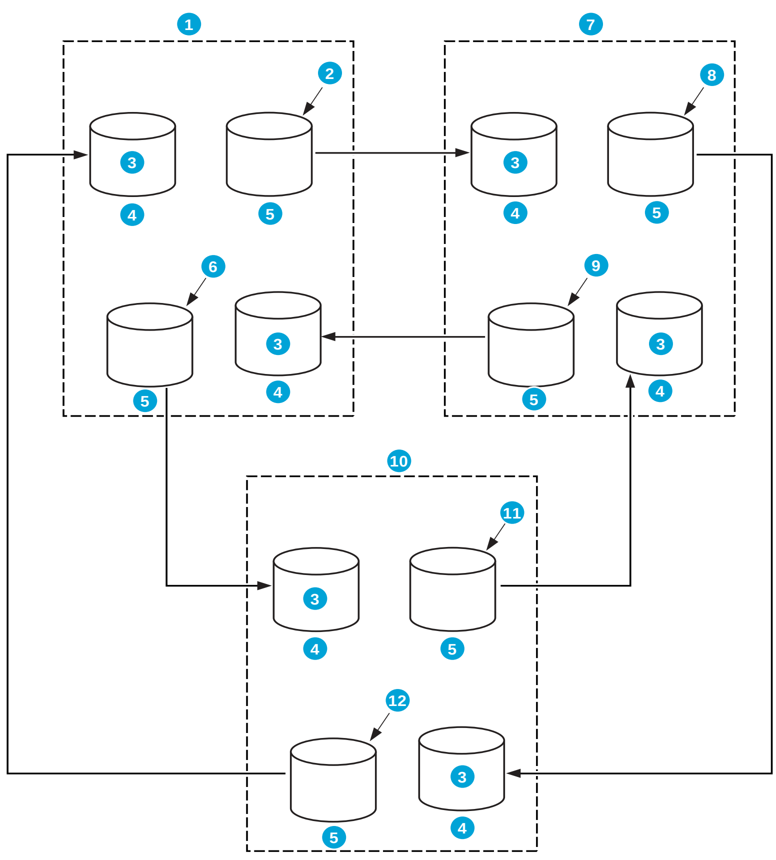
Figure 22 Cascaded replication

Bidirectional ring replication

In bidirectional ring replication, three DR groups are replicating clockwise, and another three DR groups are replicating counterclockwise. None of the six DR groups are related, other than the source disk for one DR group may be a snapclone of a destination DR group on the same array (Figure 23).

NOTE: This type of replication is impractical in all but very specialized environments.

TIP: In multiple member DR groups, you may need to quiesce the application and flush the server cache before creating a snapclone to ensure transaction level data consistency across the members of the DR group.



1	Array A	**2**	Application 1 writes
3	Copy	**4**	Destination
5	Source	**6**	Application 6 writes
7	Array B	**8**	Application 4 writes
9	Application 2 writes	**10**	Array C
11	Application 3 writes	**12**	Application 5 writes

Figure 23 Bidirectional ring relationships

5 Solution planning

This chapter describes general design considerations for planning an HP Continuous Access EVA solution. Topics include:

Operating system considerations

This section describes the operating systems supported in HP Continuous Access EVA solutions. It also describes the operating system capabilities that are available in an HP Continuous Access EVA environment. These capabilities are not always available in non-Continuous Access EVA environments.

Supported operating systems

Continuous Access EVA supports the following operating systems:

- HP-UX
- HP OpenVMS
- HP Tru64 UNIX
- IBM AIX
- Microsoft Windows NT, 2000, and 2003
- Novell Netware
- Red Hat Linux
- Sun Solaris
- SuSE Linux

For specific version and server information, see the *HP StorageWorks EVA replication compatibility reference*, which is available from the following web site:

http://h18006.www1.hp.com/products/storage/software/conaccesseva/index.html

Operating system capabilities

This section describes two operating system capabilities that are available in an HP Continuous Access EVA solution: boot from SAN and bootless failover.

Boot from SAN

An important design consideration is whether to boot servers from the SAN and, if so, whether to replicate those boot disks using Continuous Access EVA. Currently, only HP OpenVMS, HP-UX 11i, HP Tru64 UNIX, and all supported Windows operating systems support booting from the EVA-based disks. With Continuous Access EVA, it is possible to replicate those boot disks to the remote EVA for use in recovery of the server and its applications, with the data. The only restriction is that you must place high-performance files, such as page and swap files, on storage within the server, rather than on the actual EVA-based boot disk. Otherwise, there is a potentially severe performance impact to the server caused by the replication of the writes to these files.

NOTE: If you replicate a boot disk with a defined IP address, HP recommends that both sites be in the same IP subnet so that you do not need to change the address after failing over to the backup copy of the system disk. Otherwise, you must change IP addresses after a server failover, such as during the network startup.

Bootless failover

Bootless failover allows destination servers to find the new source (after failover of the storage) without rebooting the server. This capability also includes the fail back to the original source without rebooting.

NOTE: For any operating system you use, refer to the OS-specific documentation to ensure that you use compatible versions of Secure Path drivers and HBAs.

Table 5 lists which operating systems support these capabilities.

Table 5 Supported features by operating system

Legend: ✓ = supported; X = not supported

Operating system	Supported file systems	Boot from SAN (without clusters)	Boot from SAN (with clusters)	Bootless failover
HP-UX	VX and Veritas 3.5	✓	X	X
HP OpenVMS		✓	✓	✓
HP Tru64 UNIX	Advanced, UNIX (UFS)	✓	✓	✓
IBM AIX	Journal (JFS)	X	X	X
All Microsoft Windows versions[1]	NTFS	✓	✓	X
Novell Netware	Novell Storage Services (NSS); Netware volumes	X	X	X
Red Hat Linux		✓[2]	✓	✓[3]

Table 5 Supported features by operating system

Legend: ✓ = supported; X = not supported

Operating system	Supported file systems	Boot from SAN (without clusters)	Boot from SAN (with clusters)	Bootless failover
Sun	UNIX (UFS)	X	X	X
SuSE Linux	EXT2, EXT3, Reiser, and LVM	✓[4]	X	✓[5,6]

1. HP recommends that you apply all patches for security reasons.
2. Advanced Server v3.0 (32- and 64-bit) only (with and without clusters)
3. Supported in configurations using Secure Path; not supported in configurations using Qlogic Native Multipath
4. SLES 8 (32-bit and 64-bit) only
5. Supported in configurations using Secure Path; not supported in configurations using Qlogic Native Multipath
6. SLES 8 (32-bit and 64-bit) and United Linux 1 only

Windows clusters

The section describes specific issues identified with Microsoft Windows clusters.

Using multi-member DR groups

There are two restrictions you must follow when using multi-member DR groups with Microsoft Windows clusters:

- When presenting virtual disks to cluster nodes, present all members of a group to the same set of FCAs. The group cannot be split across multiple sets of FCAs. For example, if all four FCAs in a host need to access an eight-member DR group, then all of the group members can be presented to any two FCAs or to all four FCAs. However, the group cannot be split with four members presented to two FCAs and the remaining four members presented to the other two FCAs.
- When making LUN assignments, assign each shared virtual disk the same LUN number on every host. For example, if host A is assigned virtual disk 5 as LUN 3, then host B must also be assigned virtual disk 5 as LUN 3.

Using similar FCAs

Microsoft Windows clusters are supported only when all hosts use the same type of FCA. For example, if one host is using KGPSA-CA adapters, then any host in the same cluster must also use KGPSA-CA adapters.

Application considerations

Some applications do not work well in a replication environment. Consider the following:

- With a maximum of eight source-destination pairs per application, you must reconfigure some applications, such as Oracle or SAP, to reduce the number of virtual disks to eight. Some operating systems, such as HP-UX and Microsoft Windows, have a limited I/O queue depth and expect to spread I/O across many physical disks. Be especially watchful when using a high performance application on an operating system platform with limited I/O queue depth, as there may be a significant performance impact if replicating the data.
- Some applications, such as Microsoft Exchange, do not tolerate high average I/O latency and therefore may not be suitable for replication beyond a metropolitan area.

General design considerations

This section describes general design considerations for planning an HP Continuous Access EVA solution.

Failover frequency

The planned or unplanned failover of one or more DR groups should not be performed more frequently than once every 15 minutes. The planned or unplanned failover of a controller should also not be performed more frequently than once every 15 minutes.

Load balancing

HP Continuous Access EVA is most effective when the average workload (reads and writes) is applied equally to both controllers, and therefore, to both fabrics and intersite links. To obtain this balance, ensure that the utilization rate of either intersite link stays below 40%. If one link fails, the average utilization of the surviving link does not exceed 80%. Similarly, the utilization rate of a single controller on both host ports should not exceed 45% (or, at most, peak above 50%), to prevent overloading a controller, if one fails.

There are two ways to balance the workload:

- Let the hosts make the arrangements
- Prior planning when setting up the replicating virtual disks

Hosts do not share workload information with each other, therefore, you should plan to use the EVA default load balancing tools.

NOTE: HP Continuous Access EVA does not support dynamic load balancing. However, it does support static or manual load balancing.

Understanding management over distance

Another design consideration is the effect of distance and configuration size on the management of arrays that are not local to the active management server. To estimate configuration size, identify each disk group, disk drive, defined server, virtual disk, DR group, and source-destination pair as an object to be managed. As the number of objects increases, so does the time it takes to discover the objects and manage the array. Similarly, the more remote the array is from the active management server, the more time it takes to complete tasks. Combining a configuration with many objects and extreme distances can require more time to manage than is acceptable.

In Table 6, the data is based on a time limit of ten minutes to complete a management action (that is, discovery of what is in an array). If you allow more time to discover an array, then you could manage more arrays at greater distances. It may take approximately the same amount of time to discover four small configurations or two large configurations. However, if the discovery time is five minutes, plan to manage fewer arrays for any distance.

Table 6 Distance versus array manageability

Legend: ✓ = supported, recommended; — = not recommended; X = not supported

Number/size of remote EVAs	Local (< 10 km)	Metro (up to 200 km/1 ms)[1]	Regional (1–18 ms)[1]	Multiple regions (18–36 ms)[1]	Intracontinental (36–60 ms)[1]	Intercontinental (60–100 ms)[1]	Global (> 100 ms)[1]
1 small[2]	✓	✓	✓	✓	✓	✓	X
1 large[3]	✓	✓	✓	✓	✓	—	X
2 small[2]	✓	✓	✓	✓	✓	—	X
2 large[3]	✓	✓	✓	✓	—	—	X
4 small[2]	✓	✓	✓	✓	—	—	X
4 large[3]	✓	✓	—	—	—	—	X
8 small[2]	✓	✓	—	—	—	—	X
8 large[3]	✓	—	—	—	—	—	X

[1] These are one-way latencies.

[2] A small array configuration consists of 1 server using 3 DR groups and 2 copy sets per DR group, for 6 virtual disks built out of 60 disk drives in one disk group.

[3] A large array configuration consists of 10 servers using 64 DR groups and 64 copy sets, for 64 virtual disks built out of one disk group of 24 disks.

Zoning considerations

Use zoning when combining different hardware platforms, operating systems, or storage systems that are currently supported only in homogeneous SANs, and it is unknown whether there are interaction problems. Table 7 shows the zone compatibility for different platforms in a Continuous Access EVA SAN. Platforms in the same column can exist in the same SAN.

Table 7 Continuous Access EVA Platform Zoning Requirements

Zone 1	Zone 2	Zone 3	Zone 4	Zone 5
HP OpenVMS	HP OpenVMS	Linux	HP-UX	IBM AIX
HP Tru64 UNIX	HP Tru64 UNIX			
Microsoft Windows NT/2000/2003	Microsoft Windows NT/2000/ 2003			
Novell NetWare	Sun Solaris			

Controller-to-switch connections

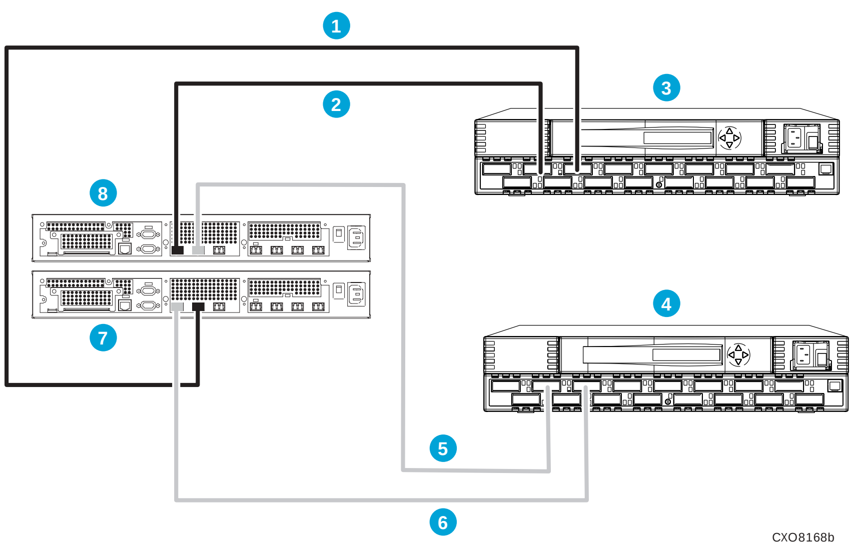
When designing your zones, consider that the controllers are to be cabled as follows.

Four fiber optic cable connections are required for each storage array. The only supported connection scheme is shown in Figure 24. Connect the fiber optic cable such that port 1 of Controller A and Controller B go to different fabrics. Connect port 2 of Controller A and Controller B to separate fabrics that are the fabric opposite from port 1 on that controller.

The following naming conventions apply to cabling:

- The storage system WWN ends with a 0.
- The Controller A, port 1 is the storage system WWN but ending with a 9.
- The Controller A, port 2 is the storage system WWN but ending with an 8.
- The Controller B, port 1 is the storage system WWN but ending with a D.
- The Controller B, port 2 is the storage system WWN but ending with a C.

A correctly cabled pair of controllers will have all ports 9 & C on one fabric and all ports 8 & D on the other. This is useful in planning the zoning of the solution.



1	B(2)	2	A(1)
3	Switch 1	4	Switch 2
5	A(2)	6	B(1)
7	Controller B	8	Controller A

Figure 24 Controller-to-switch cabling

EVA zoning recommendations

HP recommends that you zone the switches using the host WWN address for each fabric instead of the controller host port WWN. For example, the storage system host WWN is designated as 50:00:1f:e1:00:15:40:80. Cabled to this fabric are Controller A, port 2 (50:00:1f:e1:00:15:40:88) and Controller B, port 1 (50:00:1f:e1:00:15:40:8d). In Figure 25, the storage system host WWN is highlighted and the **Add Host >** button is used to place this storage system into the fabric.

NOTE: This example uses B-series switches.

On the other fabric, the storage system WWN would display as 50:00:1f:e1:00:15:40:80, with Controller A, port 1 shown as 50:00:1f:e1:00:15:40:89 and Controller B, port 2 shown as 50:00:1f:e1:00:15:40:8c. The storage system WWN is highlighted, and the **Add Host >** button is used to zone by the host WWN.

See the *HP StorageWorks SAN design reference guide* for specific EVA zoning recommendations with B-series, C-series, and M-series switches.

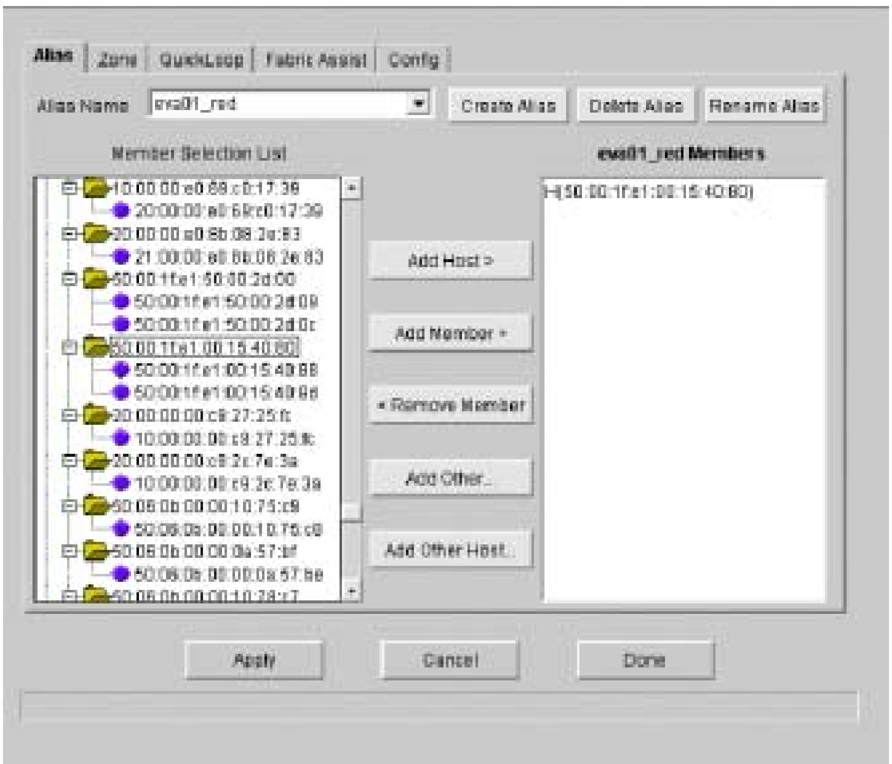


Figure 25 Example of host zoning with infrastructure switches

Glossary

active management server — *See* management server.

array — *See* virtual array and storage system.

asynchronous — A descriptive term for computing models that eliminate timing dependencies between sequential processes. In asynchronous replication, the array controller acknowledges that data has been written at the source before the data is copied at the destination. Asynchronous replication is an optional DR group property. *See also* synchronous.

bandwidth — The transmission capacity of a link or system, usually measured in bits per second.

bandwidth latency product — The measurement of the ability to buffer data and is the raw transfer speed in bytes/sec x the round-trip latency in seconds.

bidirectional — A descriptive term for an array that contains both source and destination virtual disks. This configuration allows multidirectional I/O flow among several arrays.

B-series switches — Fibre Channel core and SAN switches made by Brocade and sold by HP.

HP Continuous Access EVA — HP Continuous Access EVA is is a storage–based HP StorageWorks solution consisting of two or more storage systems performing disk–to–disk replication, along with management user interfaces that facilitate configuring, monitoring, and maintaining the replicating capabilities of the storage systems.

C-series switches — Fibre Channel switches made by Cisco and sold by HP.

data migration — Moving data to a new location or to a logical disk with a different capacity.

data mining — A process that makes data available so that undiscovered and useful information can be extracted, analyzed, or tested.

data movement — Activities such as data backup, data migration, and data distribution.

default disk group — The disk group that is created when an array is initialized. The minimum number of disks the group can contain is eight. The maximum is the number of installed disks.

destination — The targeted recipient (for example, a DR group, array, virtual disk) of replicated data. *See also* source.

disaster tolerance (DT) — The capability for rapid recovery of user data from a remote location when a significant event or disaster occurs at the local computing site. It is a special combination of high-availability technology and services that can continue the operation of critical applications in the event of a site disaster. DT systems are designed to allow applications to continue operating during the disaster recovery period.

disk group — A named group of disks selected from all available disks in an array. One or more virtual disks can be created from a disk group.

DR group — Data replication group. A named group of virtual disks selected from one or more disk groups so that they replicate to the same destination, fail over together, and preserve write order within the group.

dual fabric Two independent fabrics providing multipath connections between Fibre Channel end devices.

enabled host A host that is equipped with a replication host agent

EVA Enterprise Virtual Array, an HP StorageWorks product that consists of one or more virtual arrays. *See also* virtual arrays.

event

- A system-generated status message, resulting from a:
- - User-initiated action (for example, "suspend DR group")
- - Replication or system transaction (for example, "retrieved data for storage system")
- - Job operation (for example, "job complete")

fabric A network of Fibre Channel switches or hubs and other devices.

failover An operation that reverses replication direction so that the destination becomes the source and the source becomes the destination. Failovers can be planned or unplanned and can occur between DR groups, managed sets, fabrics or paths, and array controllers.

failsafe A descriptive term for devices that automatically assume a safe condition after a malfunction. Failsafe DR groups stop accepting host input and stop logging write history if a member of the group becomes unreachable.

general purpose server A server that runs customer applications such as file and print services. HP StorageWorks Command View EVA and HP StorageWorks Replication Solutions Manager can be used on a general purpose server in limited configurations.

home The DR group that is the preferred source in a replication relationship. By default, home is the original source, but it can be set to the destination DR group.

host A computer that runs user applications and uses (or potentially uses) one or more virtual disks that are created and presented by the array controller.

initialization A configuration step that binds the controllers together and establishes preliminary data structures on the array. Initialization also sets up the first disk group, called the default disk group, and makes the array ready for use.

managed set Selected resources grouped together for convenient management. For example, you can create a managed set to manage all DR groups whose sources reside in the same rack.

management server A server where HP StorageWorks Enterprise Virtual Array (EVA) management software is installed, including HP StorageWorks Command View EVA and HP StorageWorks Replication Solutions Manager, if used. A dedicated management server runs EVA management software exclusively. Other management servers are general purpose servers, HP ProLiant Storage Server (NAS) models, and the HP OpenView Storage Management Appliance. When there are multiple management servers in a SAN, one is active and all others are standby. The active management server actively manages the array, while the standby management server takes control of the array if there is a failure on the active management server. There is only one active management server at a time for any given management zone in a SAN.

merge The act of transferring log contents to the destination virtual disk to synchronize the source and destination.

M-series switches Fibre Channel Director and Edge switches made by McDATA and sold by HP.

near-online storage On-site storage of data on media that takes only slightly longer to access than online storage kept on high-speed disk drives.

normalization The initial full copy that occurs between source and destination virtual disks.

online storage	An allotment of storage space that is available for immediate use, such as a peripheral device that is turned on and connected to a server.
(to) present	The array controller act of making a virtual disk accessible to a host computer.
remote copy	A replica virtual disk on the destination array.
SAN	Storage area network, a network of storage devices and the initiators that store and retrieve information on those devices, including the communication infrastructure.
snapclone	A copy that begins as a fully allocated snapshot and becomes an independent virtual disk. Applies only to the HP StorageWorks EVA.
snapshot	A nearly instantaneous copy of the contents of a virtual disk created without interruption of operations on the source virtual disk. Snapshots are typically used for short-term tasks such as backups.
source (home)	A descriptive term for the virtual disk, DR group, or virtual array where an original I/O is stored before replication. *See also* destination.
standby management server	*See* management server.
Storage Management Appliance	HP OpenView Storage Management Appliance, an HP hardware-software product designed to run SAN management applications such as HP StorageWorks Command View EVA and HP StorageWorks Replication Solutions Manager.
storage system	Synonymous with virtual array. The HP StorageWorks Enterprise Virtual Array consists of one or more storage systems. *See also* virtual array.
synchronous	A descriptive term for computing models that perform tasks in chronological order without interruption. In synchronous replication, the source waits for data to be copied at the destination before acknowledging that it has been written at the source. *See also* asynchronous.
VCS	Virtual Controller Software. The software in the HP StorageWorks Enterprise Virtual Array controller. Controller software manages all aspects of array operation, including communication with HP StorageWorks Command View EVA.
virtual array	Synonymous with disk array and storage system, a group of disks in one or more disk enclosures combined with control software that presents disk storage capacity as one or more virtual disks. *See also* virtual disk.
virtual disk	Variable disk capacity that is defined and managed by the array controller and presentable to hosts as a disk.
Vraid	Techniques for configuring virtual disks to provide fault tolerance and increase performance. Vraid techniques are identified by level numbers.

Level	Redundancy	Technique
Vraid0	None	Striping
Vraid1	High	Mirroring
Vraid5	Medium	Striping and parity

Vraid0	A virtualization technique that provides no data protection. Data chunks are distributed across the disk group from which the virtual disk is created. Reading and writing to a Vraid0 virtual disk is very fast and uses available storage to the fullest, but provides no data protection (redundancy) unless there is parity.

Vraid1 A virtualization technique that provides the highest level of data protection. All data blocks are mirrored, or written twice, on separate disks. For read requests, the block can be read from either disk, which can increase performance. Mirroring requires the most storage space because twice the storage capacity must be allocated for a given amount of data.

Vraid5 A virtualization technique that uses parity striping to provide moderate data protection. For a striped virtual disk, data is broken into chunks and distributed across the disk group. If the striped virtual disk has parity, another chunk (a parity chunk) is calculated from the data chunks and written to the disks. If a data chunk becomes corrupted, the data can be reconstructed from the parity chunk and the remaining data chunks.

wavelength division multiplexing (WDM) The ability to have multiple optical signals share a single optical cable.

Index

Figures

Tables